The link between sufficient conditions by Harshman and by Kruskal for uniqueness in Candecomp/Parafac

Ten Berge, Jos M. F.; Tendeiro, Jorge N.

Published in:
Journal of Chemometrics

DOI:
10.1002/cem.1204

Research Article

Journal of CHEMOMETRICS

Received: 15 July 2008, Accepted: 9 September 2008, Published online in Wiley InterScience: 3 November 2008

(www.interscience.wiley.com) DOI: 10.1002/cem.1204

The link between sufficient conditions by Harshman and by Kruskal for uniqueness in Candecomp/Parafac[†]

Jos M. F. Ten Berge[a*] and Jorge N. Tendeiro[a]

Harshman (*UCLA Working Papers in Phonetics* 1972; 22: 111–117) has given a proof of uniqueness (identification) of Parafac solutions, when two of the three component matrices are of full column rank, and the third satisfies a few other conditions. Kruskal has given more relaxed sufficient conditions, which do not require any of the component matrices to be of full column rank. However, even when two component matrices are of full column rank, Harshman's conditions on the third matrix are still less easily satisfied than Kruskal's. The present paper bridges the gap between the two sets of conditions by utilizing the possibilities of slice mixing in Harshman's approach. It offers an alternative uniqueness theorem that is sufficiently general for all practical purposes and easy to interpret, with a proof that is easy to understand. Copyright © 2008 John Wiley & Sons, Ltd.

Keywords: uniqueness; identification; Parafac; Candecomp; three-way component analysis

1. INTRODUCTION

Richard Harshman's most important contribution to the world of multivariate data analysis has been the invention of the Parallel Factor Analysis (Parafac) method of three-way component analysis [1]. Independently, in the context of multidimensional scaling, Carroll and Chang [2] developed the Candecomp algorithm, a method for obtaining best least squares Parafac solutions. Accordingly, the method has been dubbed Candecomp/Parafac (CP). A key property of CP is the uniqueness that its components have under mild conditions. Harshman first discusses the uniqueness property in 1970, reporting a mathematical proof by Jennrich (Reference [1], pp. 61–62). This proof shows that when all three-component matrices have full column rank, alternative solutions have the same component matrices up to permutation and scale. Harshman [3] gave a more relaxed sufficient condition for uniqueness, which requires only two of the three component matrices to be of full column rank, whereas the third may have only two rows satisfying certain conditions. A further relaxation was given by Kruskal [4], who showed that even if none of the component matrices have full column rank, a CP solution may be unique. The present paper is aimed at clarifying the relation between the Harshman condition and Kruskal's in the case where two component matrices have full column rank. It will be shown that, when the possibility of taking linear combinations of the fitted data slices is fully exploited, Harshman's conditions become equivalent to Kruskal's, in cases where two of the component matrices have full column rank. In the process, we obtain an alternative uniqueness theorem that is sufficiently general for all practical purposes and easy to interpret, with a proof that is easy to understand. We start with a review of the matrix equations that define uniqueness of Parafac solutions.

2. THE EQUATIONS OF A PARAFAC SOLUTION

Let $\underline{\mathbf{X}}$ be a three-way data array of order $I \times J \times K$, containing frontal slices $\mathbf{X}_1, \ldots, \mathbf{X}_K$ of order $I \times J$. The CP method in R dimensions decomposes the slices as

$$\mathbf{X}_k = \mathbf{A}\mathbf{C}_k\mathbf{B}' + \mathbf{E}_k \tag{1}$$

where $\mathbf{A}$ is an $I \times R$ matrix, $\mathbf{B}$ is a $J \times R$ matrix, $\mathbf{C}_k$ is a diagonal matrix, containing the elements of row k of a $K \times R$ matrix $\mathbf{C}$, $k = 1, \ldots, K$, and $\mathbf{E}_k$ is a matrix of residuals. Suppose there exists an alternative solution (with the same residuals)

$$\mathbf{X}_k = \mathbf{G}\mathbf{D}_k\mathbf{H}' + \mathbf{E}_k \tag{2}$$

with $\mathbf{G}$ and $\mathbf{H}$ of the same order as $\mathbf{A}$ and $\mathbf{B}$, respectively, and $\mathbf{D}_k$ diagonal, containing the elements of row k of a $K \times R$ matrix $\mathbf{D}$, $k = 1, \ldots, K$. A solution for CP is said to be unique when, for every alternative solution of the form (2), $\mathbf{G} = \mathbf{A}\boldsymbol{\Pi}\boldsymbol{\Lambda}_1$, $\mathbf{H} = \mathbf{B}\boldsymbol{\Pi}\boldsymbol{\Lambda}_2$ and $\mathbf{D} = \mathbf{C}\boldsymbol{\Pi}\boldsymbol{\Lambda}_3$, for some permutation matrix $\boldsymbol{\Pi}$ and diagonal matrices $\boldsymbol{\Lambda}_1$, $\boldsymbol{\Lambda}_2$ and $\boldsymbol{\Lambda}_3$, with $\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_3 = \mathbf{I}_R$. That is, a solution is unique when the only changes it permits are joint permutations and rescaling of columns of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$.

* *Correspondence to: J. M. F. Ten Berge, Department of Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, Netherlands. E-mail: j.m.f.ten.berge@rug.nl*

a *J. M. F. Ten Berge, J. N. Tendeiro*
University of Groningen, The Netherlands

† *The authors are obliged to the late Richard Harshman for stimulating this research and to Alwin Stegeman for useful comments.*

Harshman [3] gave as a sufficient condition for uniqueness that two component matrices (**A** and **B**, say) have full column rank, and the third satisfies conditions of non-singularity and distinctness. That is, when elements from two rows of **C** are used to fill diagonal matrices $\mathbf{C}_1$ (non-singular) and $\mathbf{C}_2$, and all diagonal elements of $\mathbf{C}_1^{-1}\mathbf{C}_2$ are distinct, then uniqueness holds. In fact, Harshman also assumed $\mathbf{C}_2$ to be non-singular but this is not essential and will be ignored. Originally, the matrix **C** to be considered contained only two rows, but it has been well understood from the beginning that uniqueness was implied as soon as **C** contained at least one pair of rows satisfying the conditions.

It has also been clear from the beginning that Harshman's conditions were not necessary. In particular, Kruskal [4] has shown that the assumption of **A** and **B** to have full column rank is overly restrictive. In fact, neither **A** nor **B** nor **C** need to be of full column rank, as long as their so-called *k*-ranks (the *k*-rank of a matrix is the largest value *k* such that all subsets of *k* columns of the matrix are linearly independent) add up to at least $2R+2$, where *R* is the number of components. This condition is weaker than Harshman's except when $R=1$, where Harshman's condition may be satisfied but Kruskal's cannot be satisfied. Kruskal's conditions are necessary and sufficient for $R=2$ or 3, but not for other values of *R* [5].

3. WHEN A AND B HAVE FULL COLUMN RANK

In the specific situation where **A** and **B** have full column rank, the condition of Kruskal takes a particularly simple form. That is, the matrix **C** needs to have at least *k*-rank 2, which means that every pair of columns of **C** constitutes a rank-2 matrix. Equivalently, **C** has no mutually proportional columns. This condition also rules out the possibility of having a zero column in **C**. For practical applications, this is a particularly useful result because having **A** and **B** of full column rank is satisfied in nearly all applications. Unfortunately, the proof of Kruskal's general result is rather inaccessible because it relies on abstract geometrical concepts. This may well explain why textbook authors such as Smilde *et al.* [6], aiming at non-mathematicians, merely state Kruskal's condition but give a complete proof of Harshman's conditions. Although Stegeman and Sidiropoulos [7] have given a simplified version of Kruskal's proof, even their approach is still quite complicated for much of the intended readership such as chemometricians or psychometricians. At the end of the day, teachers and textbook writers face the dilemma of either presenting their students a straightforward proof for a sub-optimal result or no proof at all for the optimal result. The present paper is meant to mitigate the dilemma. It gives a straightforward proof for Kruskal's result when **A** and **B** have full column rank. In fact, Harshman's [3] proof is sharpened by taking full advantage of the possibilities of replacing the slices of the array by a set of linear combinations of these slices, a procedure which is called slice mixing.

The possibility of using slice mixing in uniqueness proofs of CP has also been considered by Leurgans *et al.* [8]. Their approach is embedded in a derivation of a closed-form CP solution, when it is known that a perfect fit solution of rank *R* exists ($R \le I$, $R \le J$). The present paper addresses the uniqueness issue directly and is, in that sense, more straightforward. Also, we use a new simplifying result (Result 2). Nevertheless, as far as uniqueness is concerned, much of the present paper is in the same spirit as that of Leurgans *et al*. From now on, the assumption that **A** and **B** have full column rank will be taken for granted.

4. RELAXING HARSHMAN'S CONDITIONS BY USING SLICE MIXING

Harshman's conditions for **C** stipulate that $\mathbf{C}_1$ and $\mathbf{C}_2$ (standing for any suitably picked pair of diagonal matrices holding elements from two rows of **C**) satisfy the condition that $\mathbf{C}_1$ is non-singular and all diagonal elements of $\mathbf{C}_1^{-1}\mathbf{C}_2$ are distinct. This poses two sorts of limitations. For one thing, when **C** has a zero element in each row, there cannot be a $\mathbf{C}_1^{-1}$. Still, Kruskal's condition of at least *k*-rank 2 may be satisfied. For example, when

$$\mathbf{C} = \begin{bmatrix} 0 & 1 & 1 \\ 2 & 0 & 6 \\ 3 & 1 & 0 \end{bmatrix} \tag{3}$$

there is no pair of proportional columns so Kruskal's condition is satisfied. However, suppose we replace our array by one in which slice $\mathbf{X}_3$ is added to slices $\mathbf{X}_1$ and $\mathbf{X}_2$, a procedure to be justified below. Then we have a CP solution with row 3 of **C** added to row 1 and row 2, so **C** is replaced by

$$\mathbf{C}^* = \begin{bmatrix} 3 & 2 & 1 \\ 5 & 1 & 6 \\ 3 & 1 & 0 \end{bmatrix} \tag{4}$$

implying that $\mathbf{C}_1$ and $\mathbf{C}_2$ end up with no zeros. The problem of zeros in each row of **C** may thus be resolved by slice mixing. However, Harshman's condition may still fail when there are no zeros. For instance, a solution with

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 6 \\ 3 & 1 & 3 \end{bmatrix} \tag{5}$$

does not satisfy Harshman's condition, because each pair of rows defines a matrix that has two proportional columns. Nevertheless, such a solution still has *k*-rank ≥ 2 for **C**. Again, adding slice 3 to the first two slices will produce a CP solution with no proportional columns in the first two rows of

$$\mathbf{C}^* = \begin{bmatrix} 4 & 2 & 4 \\ 5 & 3 & 9 \\ 3 & 1 & 3 \end{bmatrix} \tag{6}$$

and Harshman's condition for **C** will be satisfied after all.

The examples demonstrate that slice mixing may turn a CP solution that fails to satisfy Harshman's condition into one that does satisfy it. By itself, slice mixing is an operation that is not allowed in CP, because it changes the data and the new CP solution (with the same slice mix applied to **C**) may no longer be the optimal CP solution for the transformed array. Still, the slice mix in no way affects the properties that determine whether or not a CP solution is unique (Reference [5], p. 401). Specifically, we may pre-multiply **A**, **B** and **C** by non-singular matrices $\mathbf{N}_A$, $\mathbf{N}_B$ and $\mathbf{N}_C$, and transform the fitted data array accordingly, (e.g. when **A** is pre-multiplied by $\mathbf{N}_A$, so are the frontal slices of $\mathbf{AC}_k\mathbf{B}'$, and so on) without affecting (non-)uniqueness. For instance, when $\mathbf{AC}_k\mathbf{B}' = \mathbf{GD}_k\mathbf{H}'$, $k=1,\ldots,K$, implies $\mathbf{G} = \mathbf{A\Pi\Lambda}_1$, as defined below

(2), then $\mathbf{N}_A\mathbf{AC}_k\mathbf{B}' = \mathbf{N}_A\mathbf{GD}_k\mathbf{H}'$, $k = 1,\ldots,K$, for non-singular $\mathbf{N}_A$, also implies $\mathbf{G} = \mathbf{A}\boldsymbol{\Pi}\boldsymbol{\Lambda}_1$, and likewise for $\mathbf{N}_B$ and $\mathbf{N}_C$.

Accordingly, we can relax Harshman's condition for **C** by considering not just every pair of rows in **C**, but all possible pairs of rows that may arise when **C** is pre-multiplied by a non-singular matrix $\mathbf{N}_C$. That is, instead of considering all pairs of rows of **C**, we may consider all pairs of vectors in the row space of **C**. This approach allows us to show that when **A** and **B** have full column rank and **C** has no mutually proportional columns, a CP solution is unique (Result 3). However, two preliminary results will be considered first.

Result 1. For any finite set of non-zero vectors $\mathbf{x}_1,\ldots,\mathbf{x}_p$, there exist vectors that are orthogonal to none of them.

Proof. Trivial. In fact, any vector **y** that is randomly generated from a continuous distribution will be non-orthogonal to $\mathbf{x}_1,\ldots,\mathbf{x}_p$ with probability 1.

An immediate implication is that the row space of **C** always contains a vector with no zeros, when **C** has no zero column, a property to be used in Result 3 below. But the result also plays a role in the following:

Result 2. The row space of **C** contains a pair of vectors $\mathbf{x}'\mathbf{C}$ and $\mathbf{y}'\mathbf{C}$ satisfying the distinctness condition if and only if **C** has no mutually proportional columns.

Proof. When at least two columns of **C** are proportional, every $2 \times R$ matrix $\begin{bmatrix}\mathbf{x}'\mathbf{C}\\ \mathbf{y}'\mathbf{C}\end{bmatrix}$ will have the same proportionality, and distinctness is impossible. Conversely, suppose there is no pair of mutually proportional columns. Then each $\mathbf{W}_{ij}$, defined as $\mathbf{c}_i\mathbf{c}_j' - \mathbf{c}_j\mathbf{c}_i'$, where $\mathbf{c}_i$ and $\mathbf{c}_j$ are columns of **C** ($i < j$), is a non-zero matrix ($\mathbf{c}_i\mathbf{c}'_j = \mathbf{c}_j\mathbf{c}'_i$ would imply $\mathbf{c}_i$ to be proportional to $\mathbf{c}_j$). Let **V** be a $K \times .5R(R-1)$ matrix holding one non-zero column from each $\mathbf{W}_{ij}$ ($i < j$). Pick a vector **x** that is non-orthogonal to all columns of **V** and to all columns of **C**, see Result 1. Then the elements of $\mathbf{x}'\mathbf{C}$ define a non-singular diagonal matrix, and the vectors $\mathbf{x}'\mathbf{W}_{ij}$ are all non-zero. Pick a vector **y** non-orthogonal to these vectors, see again Result 1. Then $\mathbf{x}'(\mathbf{c}_i\mathbf{c}'_j - \mathbf{c}_j\mathbf{c}'_i)\mathbf{y} \neq 0$ for all $i < j$. This means that $(\mathbf{x}'\mathbf{c}_i)(\mathbf{y}'\mathbf{c}_j) - (\mathbf{x}'\mathbf{c}_j)(\mathbf{y}'\mathbf{c}_i) \neq 0$ for all $i < j$, so every 2×2 sub-matrix of the $2 \times R$ matrix holding $\mathbf{x}'\mathbf{C}$ and $\mathbf{y}'\mathbf{C}$ as rows has non-zero determinant. It follows that, when **C** is pre-multiplied by a non-singular matrix $\mathbf{N}_C$ with $\mathbf{x}'$ and $\mathbf{y}'$ as first two rows, the two diagonal matrices holding elements from the first two rows of $\mathbf{N}_C\mathbf{C}$ satisfy the distinctness condition.

We are now in a position to present a relaxed version of Harshman's condition which is equivalent to Kruskal's condition in the case under consideration (with **A** and **B** of full column rank).

Result 3 (Relaxed Harshman condition). When **A** and **B** have full column rank and **C** has no mutually proportional columns, then a CP solution is unique.

Proof. Find a vector **x** not orthogonal to any column of **V** and of **C** (see Result 2), and find a vector **y** not orthogonal to any of the vectors $\mathbf{x}'\mathbf{W}_{ij}$. Let $\mathbf{C}_1$ and $\mathbf{C}_2$ be diagonal matrices holding the elements of $\mathbf{x}'\mathbf{C}$ and $\mathbf{y}'\mathbf{C}$. Then $\mathbf{C}_1$ is non-singular and $\mathbf{C}_1$ and $\mathbf{C}_2$ satisfy the distinctness condition. From (1) and (2) we have $\mathbf{A} = \mathbf{GD}_1\mathbf{H}'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{C}_1^{-1}$ so **A** spans the same column space as **G**. Likewise, **B** spans the same column space as **H**. Hence, $\mathbf{G} = \mathbf{AS}$ and $\mathbf{H} = \mathbf{BT}$ for non-singular matrices **S** and **T**. Rewriting (2) as $\mathbf{X}_k = \mathbf{ASD}_k\mathbf{T}'\mathbf{B}' + \mathbf{E}_k$ and using (1) yields $\mathbf{AC}_k\mathbf{B}' = \mathbf{ASD}_k\mathbf{T}'\mathbf{B}'$. Removing **A** and **B** yields

$$\mathbf{C}_k = \mathbf{SD}_k\mathbf{T}' \tag{7}$$

$k = 1,\ldots,K$, with **S** and **T** non-singular. From (7) we have $\mathbf{C}_2\mathbf{C}_1^{-1} = \mathbf{SD}_2\mathbf{T}'(\mathbf{T}')^{-1}\mathbf{D}_1^{-1}\mathbf{S}^{-1} = \mathbf{SD}_2\mathbf{D}_1^{-1}\mathbf{S}^{-1}$, or, equivalently, $\mathbf{C}_2\mathbf{C}_1^{-1}\mathbf{S} = \mathbf{SD}_2\mathbf{D}_1^{-1}$, and that $\mathbf{C}_2\mathbf{C}_1^{-1} = \mathbf{TD}_2\mathbf{S}'(\mathbf{S}')^{-1}\mathbf{D}_1^{-1}\mathbf{T}^{-1} = \mathbf{TD}_2\mathbf{D}_1^{-1}\mathbf{T}^{-1}$, or, equivalently, $\mathbf{C}_2\mathbf{C}_1^{-1}\mathbf{T} = \mathbf{TD}_2\mathbf{D}_1^{-1}$. It can be seen that **S** and **T** hold eigenvectors of $\mathbf{C}_2\mathbf{C}_1^{-1}$.

At this point, standard matrix algebra can be used. When all diagonal elements of a diagonal matrix are distinct, so are its eigenvalues, and the associated eigenvectors are columns of the identity matrix, up to permutation and scale. Hence, $\mathbf{S} = \boldsymbol{\Pi}_1\boldsymbol{\Lambda}_1$, with $\boldsymbol{\Pi}_1$ a permutation matrix and $\boldsymbol{\Lambda}_1$ diagonal. In the same vein, $\mathbf{T} = \boldsymbol{\Pi}_2\boldsymbol{\Lambda}_2$ with $\boldsymbol{\Pi}_2$ a permutation matrix and $\boldsymbol{\Lambda}_2$ diagonal. From (7), written as $\mathbf{C}_k = \boldsymbol{\Pi}_1\boldsymbol{\Lambda}_1\mathbf{D}_k\boldsymbol{\Lambda}_2\boldsymbol{\Pi}'_2$, and the fact that $\mathbf{C}_1$ is non-singular, it follows that $\boldsymbol{\Pi}_1 = \boldsymbol{\Pi}_2 = \boldsymbol{\Pi}$, so $\boldsymbol{\Pi}'\mathbf{C}_k\boldsymbol{\Pi} = \boldsymbol{\Lambda}_1\mathbf{D}_k\boldsymbol{\Lambda}_2$, $k = 1,\ldots,K$. Hence $\mathbf{C}\boldsymbol{\Pi} = \mathbf{D}\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_2$, so $\mathbf{D} = \mathbf{C}\boldsymbol{\Pi}\boldsymbol{\Lambda}_1^{-1}\boldsymbol{\Lambda}_2^{-1} \equiv \mathbf{C}\boldsymbol{\Pi}\boldsymbol{\Lambda}_3$, with $\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_3 = \mathbf{I}_R$.

5. DISCUSSION

We have relaxed Harshman's approach to obtain sufficient conditions for uniqueness in CP, by taking advantage of the possibilities of slice-mixing. Compared to Kruskal's condition, Harshman's approach is more accessible because it relies on standard tools of matrix algebra. This may be convenient for pedagogical purposes. From a mathematical perspective, however, Kruskal's condition remains more powerful because it also handles cases where neither **A**, nor **B**, nor **C** has full column rank.

The above proof shows that Harshman's condition on **C**, once extended with the possibilities of slice mixing, is equivalent to having *k*-rank 2 at least for **C**. It is well known that when **A** and **B** have full column ranks, this condition on **C** is also necessary for uniqueness. For instance, Leurgans *et al*. [8] have shown how to obtain alternative solutions when two columns in **C** are proportional.

REFERENCES

1. Harshman RA. Foundations of the Parafac procedure: models and conditions for an 'explanatory' multimodal factor analysis. *UCLA Working Papers in Phonetics* 1970; **16**: 1–84.
2. Carroll JD, Chang JJ. Analysis of individual differences in multidimensional scaling via an n-way generalization of 'Eckart-Young' decomposition. *Psychometrika* 1970; **35**: 283–319.
3. Harshman RA. Determination and proof of minimum uniqueness conditions for Parafac1. *UCLA Working Papers in Phonetics* 1972; **22**: 111–117.
4. Kruskal JB. Three-way arrays: rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics. *Linear Algebra & Applications* 1977; **18**: 95–138.
5. Ten Berge JMF, Sidiropoulos ND. On uniqueness in Candecomp/Parafac. *Psychometrika* 2002; **67**: 399–409.
6. Smilde AK, Bro R, Geladi P. *Multi-way analysis with applications in the chemical sciences*. Wiley: New York, 2004.
7. Stegeman AW, Sidiropoulos ND. On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition. *Linear Algebra & Applications* 2007; **420**: 540–552.
8. Leurgans SE, Ross RT, Abel RB. A decomposition for three-way arrays. *SIAM J Matrix Anal Appl* 1993; **14**: 1064–1083.

